DOCUMENT RESUME

ED 217 711 FL 012 977

AUTHOR Stansfield, Charles
TITLE Reliability and Validity of the Secondary Level English Proficiency Test.
PUB DATE 82
NOTE 29p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *English (Second Language); *Language Proficiency; *Language Tests; Listening Comprehension; Multiple Choice Tests; Reading Tests; Secondary Education; Testing; *Test Reliability; *Test Validity
IDENTIFIERS *Secondary Level English Proficiency Test

ABSTRACT

The secondary level English proficiency (SLEP) test is a group administered 150 item multiple-test of English language proficiency that includes two subscores and eight different item types. It is designed to assess a foreign student's readiness for English medium instruction at the secondary level. This paper reports on two studies which were conducted during field testing of the instrument. The results indicate high reliability for the total test (.96) and for each subtest (.94 and .93). A validity study involved analysis of test scores and demographic data for U.S. public school students. The data grouped students according to citizenship status, length of time in school, length of time in the United States, length of English study within and outside the United States, and grade. The results indicate consistent growth in the expected direction for subgroups established for each variable. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED217711

RELIABILITY AND VALIDITY OF THE SECONDARY LEVEL ENGLISH PROFICIENCY TEST

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Charles Stansfield

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Charles Stansfield

Educational Testing Service

Princeton, New Jersey

FL012977

The Reliability and Validity of the Secondary level English Proficiency Test

INTRODUCTION

This paper is a report to the profession on a new test English as a second language, the Secondary Level English Proficiency (SLEP) test. The report is intended to meet two objectives: to provide information that might be of use to item writers and others interested in various techniques for assessing language skills; and to provide additional information, not available elsewhere, which might be of use in judging the overall validity of the test. Regarding the second objective, we view the establishment of validity as the marshalling over a period of time of evidence which would either support or would not support the use of a test for a particular function or for a particular type of examinee. Thus, validity is basically a judgmental or inferential matter, rather than a purely empirical one.

The SLEP is a 150 item, four-option multiple-choice test of English language proficiency. The test provides a total score and diagnostic subscores that measure ability in two primary areas: understanding spoken English and understanding written English. Henceforth, we will refer to the sections that measure these areas as the "listening" and "reading" sections of the test. Each section contains 75 items. Including the time required for listening to the directions and doing sample items, Section One, listening, lasts 40 minutes, and Section Two, reading, lasts 45 minutes. The total time required to administer both sections of the test is one hour and twenty-five minutes.

The SLEP was developed with support provided by the TOEFL Policy Council, as a secondary school version of the Test of English as a Foreign Language. Published by Educational Testing Service (ETS), SLEP is designed for use as a selection or admissions instrument by private secondary schools, or as a placement instrument by public secondary schools.(1)

HISTORY

The history of the SLEP dates back to the mid 1970s, when ETS received frequent inquiries from private secondary schools in the United States and abroad regarding the development of a lower level version of the TOEFL. In response to the interest expressed, in 1976 TOEFL staff sent a questionnaire to 500 private domestic and overseas secondary schools. The questionnaire sought information on the schools' need for an English language test for selection and placement purposes, the English language screening procedures currently used in the admissions process, and the degree of interest in a lower level TOEFL. Over 60% of the schools returning the questionnaire indicated support for the development of a secure English as a second language proficiency test for secondary schools. Subsequent contacts with officers of the National Association of Independent Schools (NAIS) and other knowledgeable persons indicated that the foreign student population in private secondary schools had doubled between 1974 and 1978, and, as in universities, the population was continuing to grow. By 1978, over 25,000 foreign students were enrolled in private secondary schools belonging to the NAIS organization.

In 1977, ETS staff accessed data on the date of birth of 1976 TOEFL registrants. It was found that 1,984 students between ages 12 and 16 had taken the TOEFL. These students were probably too young to enter

college, and therefore it was assumed that their reason for taking the TOEFL was related to a desire to be admitted to private secondary schools.

In October 1978, ETS invited representatives of public and private secondary schools enrolling a large number of international students to a meeting at which the possibility of developing a lower level TOEFL was discussed. Strong support was expressed for such a measure. Feedback indicated that the TOEFL, which emphasizes college level academic English, was too difficult for this group of students and not adequately focussed on the kind of language they encounter. The indication was that listening and reading were the communicative skills that should be assessed. Although some support was expressed for an actual measure of writing, it was assumed that this skill was sufficiently related to reading proficiency that a separate measure would not be necessary. The following month, the TOEFL Policy Council approved a proposal to develop the SLEP.

TEST DEVELOPMENT

Preliminary SLEP item types were developed by ETS staff and submitted to members of a Committee of Examiners composed of six secondary school teachers of English as a second language. During 1979 meetings were held to review potential item types and discuss test specifications. Subsequently, a fourteen page set of test specifications was developed. Early in the test development process, it was decided to use a multiple-choice item format. This format helps insure score reliability through the standardization of administration procedures and also eliminates the need to rely on the subjective judgment of raters. The choice of material for the test was based on an analysis of actual textbooks designed for use in American classrooms in grades 7-11. Regarding the social context of

items, the committee decided to present situations representative of those encountered by students in American secondary schools. This design decision is particularly evident in the conversations used to test listening comprehension.

Eventually, almost 1,500 questions were written and reviewed by ETS test development staff and by the secondary school ESL teachers. Following review, sixteen different pretests were administered to students in order to gather data on item performance. Over 6,000 students in 30 secondary schools representing 12 countries in North America, Central America, South America, Africa, Asia, Europe and the Middle East, took a form of the pretest. Subsequently, it was determined that 84 percent of the pretested items could be used in operational forms of the test. Two such forms have since been assembled. The second form is equated to the first through common item equating. Examinee score and item response data from the first form are examined in depth below.

TEST CONTENT

In addition to the history stated already, each of the eight item types selected for inclusion in the specifications and on the test will be described briefly in order to provide information that can be used to judge the content validity of the test.

Section One, the listening section, contains four parts, each of which has one item type. Part One requires the student to comprehend and correctly identify a sentence describing a single picture stimulus. The student hears four sentences and marks the letter of the sentence that correctly describes the picture. The SLEP contains 25 such listening comprehension items, dealing with correct recognition of minimal pair

contrasts, juncture, stress, sound clusters, tense voice, prepositions, and vocabulary.

Part Two consists of 12 items based on a map drawn in the test booklet. The map represents the downtown area of a small town, including buildings, parks, street names, etc., and depicts four cars labelled A, B, C, and D. After listening to a brief conversation between two people, the student must decide in which car the conversation occurred. Each conversation discusses how the occupants of the car will get from where they presently are to where they want to go. (e.g., "I'd like to very much. If we continue on Mackerel to the circle and go around to Salmon, we can park on Cod Lane.") This part assesses an integrated variety of linguistic, cultural, and pragmatic concepts. These include directions (e.g., compass points -- north, south, east, west; turns -- right, left, straight; street relationships), recognition of building names and association of appropriate vocabulary with the building (e.g., snack - restaurant), distances, and time. Map items of listening comprehension are essentially pure items, since very little reading is involved.

Part Three of the listening subtest consists of 28 items based on extended conversations. These conversations, representing typical secondary school situations, were recorded by American high school students. For each recorded question, the student must choose one of the four answers printed in the test book. The conversations take place in various parts of the school (e.g., cafeteria, library, study hall, counselor's office, nurse's office, etc.), and deal with events that typically occur in each location (e.g., gymnasium - pep rally). The conversations, may also deal with extracurricular activities, such as car washes, bake sales, yearbook,

or sports. Typical high school academic subjects, such as civics, geometry, and current affairs, or non-academic matters such as school closings, holidays, and vacations may also be the topics of these conversations. As can be seen, this and other parts of the test are in no sense "culture-free." Rather, a deliberate attempt is made to link the language tested to an appropriate cultural milieu. Thus, the SLEP is a language proficiency measure based on the language that is likely to be encountered by a student attending high school in the United States or an American type high school overseas.

Part Four consists of 20 items involving what we have called a multiple-choice dictation. This item type obviates the complaint (Lado, 1961) that the dictation tests too many elements at once and overcomes the problem of subjective grading of responses. The student must match one of four sentences printed in the test book with a sentence heard on tape. Many of the sentences are the type of utterance the student is likely to hear from the teacher or classmates (e.g., "Shouldn't you do your reading assignment before answering the questions?"). The distractors emphasize structural variations rather than phonological problems. All distractors are grammatically correct and none is merely a rephrasing of the keyed distractor. We have called this a dictation because it functions as a dictation psycholinguistically. The student must retain the complete thought in short term memory, while he constructs how it should be written. This process is similar to writing the sentence without confusing or misinterpreting what was dictated.

Parts Five through Eight are found in Section Two, which is the reading comprehension subtest. Part Five consists of twelve items based

on a single cartoon. The cartoon depicts a specific event, such as a school closing due to a heavy snow storm, and the reaction of several members of a family to that event. Each family member is labelled with the letter A, B, C, or D in the cartoon. The examinee reads a series of stimulus utterances each of which is composed of two or three sentences, and indicates which member of the family probably made each utterance based on the situation depicted. This part tests the reading of short passages that describe a common situation using everyday vocabulary.

The student sees four line drawings for each item in Part Six. After reading a single sentence, the student must indicate which drawing the sentence describes. The items in this section test prepositions, adjectives (e.g., words indicating quantity and size, as well as the comparative and superlative forms), adverbs, pronouns (case), and other words that indicate the relationship between the people or objects portrayed.

Part Seven consists of 40 items based on three multiple-choice cloze passages testing a wide variety of grammatical and lexical elements. In fourteen of these forty items, the student answers a series of reading comprehension questions after the passage based on the information contained in the cloze passage. Thus, the cloze is very efficiently used as both a test in itself, and the stimulus paragraph about which a series of other questions are asked.

Part eight consists of eight reading comprehension items based on a 140 word literary passage from a high school textbook. The examinee must comprehend and recall details of the passage and make inferences as to the main idea, tone, relationships between characters, events, the author's purpose, etc.

Subsequent to the assembly of the first operational form, the SLEP was administered to foreign students enrolled in American public schools and to non-native English speaking students enrolled in private schools overseas in order to gather data on its reliability and validity. The information presented in the remainder of this paper is based on test performance and demographic data produced by those two different populations.

TEST ANALYSIS

In the first study reported here, the SLEP was offered as an alternative to the TOEFL to foreign secondary school students applying for admission to American private secondary schools. It was given at secure international administrations in January and May, 1980, to a total of 310 students in 25 countries. From this data, a standard scale was developed for equating scores across forms, and a statistical analysis of the test's overall performance was carried out. The standard scale is based on the T-score, which has a mean of 50 and a standard deviation of 10. Because the raw scores gathered in the first two international administrations exhibit a ceiling effect, scaled scores range from a low of 20 to a high of 67.

Insert Table I about here.

Table I portrays the raw score distributions for each of the eight parts of the test. Middle difficulty is the midpoint between the maximum possible score and the score that would be expected if each item were answered at random. For four-option tests like the SLEP, middle difficulty corresponds to 62.5 percent of the maximum possible score. The table shows that all but two of the parts are easier than middle difficulty.

Part Two, the multiple-choice dictation, was the easiest part of the test. The average performance on this part for all students tested was 90% correct. Thus, it appears that such items are generally quite easy, although their difficulty can be increased through the use of good distractors.(2)

Almost equal in facility (89% correct) were [illegible] items based on a cartoon. One could argue that performance on this part is related to familiarity with American family life which it portrays. The data presented in Table I was obtained from students living outside the United States. The fact that they performed so well on it indicates that the cultural referents in the cartoon do not interfere with the ability to answer the questions. The items based on pictures (i.e., a single picture per item, four pictures per item, and items based on a town map), were next in level of difficulty, and were also relatively easy.

The three most difficult item types were those that did not contain pictures (i.e., comprehension of extended conversation, cloze, and comprehension of a literary passage). Performance on such integrative items involves global understanding of the context as well as recognition of discrete elements of language. Since the SLEP is one of the first standardized foreign language tests to use multiple-choice cloze items, it is interesting to observe how examinees performed on these items.(3) As can be seen in Table I, these items were of middle difficulty (62% correct), which is the ideal difficulty level for maximum discrimination.

RELIABILITY

The last row of Table I shows the KR-20 index of internal consistency reliability. This represents parallel-form reliability based on the

inter-item consistency for each part. Thus, inter-item consistency serves as a surrogate measure of parallel-form reliability, thereby eliminating the need to administer and correlate two different forms of the test. As can be seen, the two experimental item types, multiple-choice dictation and multiple-choice cloze, showed the highest reliabilities. However, the cloze procedure is the more efficient. Ninety-three percent of pretested cloze items were usable, whereas only 63% of the pretested dictation items were usable. For the SLEP, a usable item was one whose biserial correlation with total score for the section in which it appears is greater than .30. Such items are considered minimally efficient discriminators of language skills. A greater proportion of cloze items, as compared to the other item types written by staff, demonstrated this criterion of discrimination power. This suggests that in the future test authors would do well to consider cloze items for inclusion on standardized second language tests.

Insert Table II about here.

Table II depicts some descriptive statistics by section for the same sample of adolescent students. The data indicate that SLEP total score and section scores are highly reliable. This is due in part to the previously mentioned decision, made at the test design stage, to utilize multiple-choice item formats only. It is also due to the test's length and item discrimination power. The items included on the final form of the test discriminate well. The mean biserial correlation with section score for items in the listening section is .61. For items in the reading section it is .55. In spite of its favorable item discrimination power, the test

remains relatively easy and is not significantly speeded. On the reading comprehension section, which is not paced, 87% of the students finished the test, and the mean number of items not reached was 1.0.

CONVERGENT AND DIVERGENT VALIDITY

Table III shows the observed correlations between part scores and section scores for the eight parts of the test. Intercorrelations among the four parts in section I range from .55 to .72 and show a mean value of about .63. Thus, it appears that the four parts are generally measuring different but related aspects of language. The dictation has relatively lower correlations with Parts Three and Four of the section than it does with part one. This could be related to the use of single sentences in Parts One and Two, as opposed to the use of multiple sentence conversations in Parts Three and Four. Each part correlates well with the section score.(4)

Insert Table III about here.

The intercorrelations among the four parts in Section II were more disparate ranging from .36 to .77. Part Five, containing items based on a cartoon, shows the lowest intercorrelations with the other parts and with the section score. In part this is due to the fact that this part is less reliable than most of the others. Also, the fact that it was quite easy ($\bar{X}$ = 89% of total), limiting the variability among scores, probably contributed to the lower correlations between it and other parts of section II. Again, the excellent performance of the cloze passage is noteworthy. Its observed correlation with the section score was

very high (.90). Such an impressive outcome suggests that a multiple-choice cloze shows considerable promise as an overall measure of reading proficiency.

The correlations in Table III provide support for the conclusion that the eight parts of the test measure different but moderately interrelated aspects of the skill being assessed by the test. In total they yield a correlation of .78 between sections, which, given the high reliability of each, means that each section is measuring aspects of language acquisition that are closely related but not identical. Thus, each contributes some unique variance to the total score, except for the cloze task, which for this sample functioned as a near perfect predictor of the reading comprehension score.

CRITERION-RELATED VALIDITY

Before the SLEP could be used in American public schools, it was necessary to obtain data on the performance of various groups of nonnative English speakers enrolled in different public school programs. The data would make it possible to compare an individual student's performance with the performance of other students with similar background characteristics, and to use this information in determining appropriate placement in a remedial or mainstream program. In order to accomplish this, a free administration of the SLEP was provided to a large number of students and at the same time basic information on their background and current educational placement was gathered. The following procedures were employed.

In August, 1980, one thousand secondary schools located in or near large metropolitan areas were randomly selected to receive a letter inviting each to participate in a validation study. Free test materials,

scoring services and school score rosters were offered as an incentive to participate. Sixty-eight schools representing 20 states responded affirmatively to the invitation. The SLEP was administered by local school personnel during the months of September and October according to standardized procedures described in the SLEP Supervisors Manual. Individual shipments of test materials also contained a supply of questionnaires to be filled out by students (See Appendix A). The questionnaire requested student responses to six questions regarding their visa status, enrollment in remedial or mainstream programs, and exposure to English. Of the 1,744 students who took the test, 1,239 returned the questionnaire.

Insert Table IV about here.

Table IV depicts student responses to the question, "In which type of program are you currently enrolled?" Student responses indicate consistent improvement in SLEP scores according to the degree of identified remedial instruction. Students receiving part-time remedial instruction performed better than those receiving full-time remedial instruction, and students who were mainstreamed performed considerably better than those who were receiving partial remedial assistance via special programs. It should be remembered that the data presented here do not represent a single program but rather a total of 68 programs. No attempt was made to insure the comparability of programs within each type. It is highly probable that programs vary considerably from school to school. Indeed, this is indicated by the standard deviation, which is larger than might be expected if the

programs were similar and homogeneous. Also, it should be remembered that most schools do not simultaneously offer both bilingual education and English as a second language instruction, and that no information was collected as to the number and type of programs available in each school. Therefore, it should not be concluded that students enrolled in ESL programs are more proficient than students enrolled in programs of bilingual education. Such a hypothesis was not part of this project.

The comparative score data in Table IV are also presented in the Manual for Administering SLEP along with comparative score data on other background variables included on the SLEP answer sheet and on the questionnaire. Similar tables include means and standard deviations by age, grade, visa status, sex, and exposure to English. This information will not be represented here. Rather, we will use the data to gather some evidence for the criterion-related validity of the SLEP. It was understood at the start of this process that the relationship between scores and criterion variables would be limited due to numerous sources of unreliability, some of which have already been mentioned.

With these cautions in mind, the product moment correlations between SLEP scores and treatments were determined by recoding a full-time treatment as a "1," a part-time treatment as a "2," and a regular class placement as a "3." The observed correlations between score and program assignment were moderate (.33 for ESL programs and .57 for bilingual education programs). These correlations are attenuated by numerous factors, including the unequal number of subjects sampled, the small number of placement categories, the lack of consistency in programs across schools (part-time could mean 20 minutes per week or three hours per day), the lack of

reliability and validity in the placement procedures actually employed (placement by surname, etc.), and the fact that the questionnaires were filled out by students with limited English proficiency. When viewed in this context, the predictive validity of the SLEP tentatively appears to be good. Controlled studies within individual schools, in which SLEP is administered concomitantly with other previously validated placement procedures, will be necessary in order to establish a more accurate estimate of predictive validity.

Table V portrays the relationship between student test scores and responses to other questions in the questionnaire. Because not all persons responded to each question, the N for each criterion variable is indicated. The data indicate that all variables are significantly related to SLEP scores, although the strength of the relationship varies according to the characteristic assessed.

Insert Table V about here.

Years of English study (1, 2, 3, 4 or more) showed the strongest correlation with SLEP score. This information was indicated by students on the SLEP answer sheet. This includes instruction both within and outside of the United States. The correlation with total score was .41. The number of years of English study within the United States, as indicated in three response categories, was more strongly correlated with SLEP score than was the amount of English study outside of or prior to coming to the United States. This suggests that for a cross-section of nonnative students enrolled in United States' public schools, the formal classroom

instruction received before arriving here plays only a small role in explaining actual overall language proficiency. Time spent studying English subsequent to arrival is a more effective determiner of actual or current proficiency as indicated by SLEP score ($r = .34$), as is time spent in the United States ($r = .35$). The length of enrollment in the current school is also related to SLEP score ($r = .25$), although not as strongly as other variables that are associated with time spent in the United States.

In general, the data indicate that the listening comprehension score correlates slightly more strongly with each criterion variable than does the reading score. This is not surprising since listening comprehension involves skill in understanding spoken English. This skill is more easily acquired in an immersion environment, such as that represented by residence and study within the United States, than in a foreign language environment. Thus, listening comprehension is more sensitive to the variables assessed here. In this respect, and given the high reliability (.94) of this subscore, it may be permissible in some contexts to utilize a short-form of the SLEP consisting of the listening comprehension section only. On the other hand, such a recommendation should be considered with caution since the criterion variables utilized here are not those associated with success in school (e.g., grade point average).

It should be cautioned that all of the correlations presented here are attenuated, since the number of possible responses was restricted to between three and six. It is likely that SLEP scores would show higher correlations with residence and schooling in the United States if a greater number of response categories were available. Since only a few

categories were used here, these low to moderate correlations must stand as modest indices of the criterion-related validity of the SLEP.

This article has attempted to provide additional data which could be used to evaluate the reliability and validity of the SLEP, and data on the performance of the various item-types used. It presents information not reported in the test manual and further analyses of existing data. At this time it is fair to conclude that SLEP scores, both part and total, are highly reliable. The content validity of the test is good, particularly for English as a second language students enrolled in grades seven through twelve. The construct validity also appears to be good, since the analysis indicates that the parts and sections are measuring different but interrelated aspects of language proficiency. While favorable evidence of criterion-related validity was presented, additional research is needed in this area. This research should include studies at the district level of the SLEP's ability to predict teacher placements, as well as placements determined by local instruments or procedures whose validity has already been established. In addition, SLEP scores should be examined for their relationship to grades earned in mainstream classrooms, and with scores on local and national achievement tests. Finally, the SLEP scores of native speakers at different grade levels should be determined in order to gain a conceptual understanding of a "native speaker level" of performance, and of the language skills differential that may still exist between nonnative and native English speaking students at any given grade level.

NOTES

(1) I wish to acknowledge the contributions to this report of my colleagues Francean Meredith, who supervised the test development process; Ann Angell, who performed the statistical analysis of the data obtained from the first two administrations of the test and obtained the part-section correlations; Nancy Turner, who performed the correlational analysis of the public school data; and Paul Angelis, who provided overall direction for the test project during its formative years. Ann Angell also made helpful comments on earlier versions of the manuscript, as did Gordon Hale, Gay MacQueen, and Russell Webster.

(2) DiFiore (1980) analyzed 56 SLEP pretest dictation items provided by ETS. The items that functioned best had distractors that resembled their keys in four areas: word position, syntax, semantics, and phonology. Good distractors used the same word as the key at the beginning and at the end of the sentence. They also used parallel syntactic structures. The following pretest item exemplifies this phenomenon.

A. I wish Dr. Miller could tell me what to do.

*B. I wish I could tell you where Dr. Miller is.

C. I hope you can tell me who Dr. Miller is.

D. I thought I knew where Dr. Miller is.

The correct response is B, and the best distractor is C. Each sentence consists of a main clause and two dependent clauses. However, distractor A ends with an infinitive in the third clause, and distractor D lacks a modal and a direct object in the second

clause. Thus, their syntactic structure differs from that of the keyed option. Semantic similarities refer to the use of similar yet different elements within sentences, such as in adverbial phrases indicating destination. Overall phonological similarity, the degree to which the options sound alike, is the most pervasive element in creating good distractors.

(3) During item analysis, a printing error on one of the reading comprehension items based on a cloze passage was discovered in the test booklet. Although subsequently corrected, examinee responses to this item were not counted in the test analysis data reported here.

(4) These correlations have been corrected for spuriousness. Correlations between scores having items in common, such as the part score with the total score, are spuriously high. A statistical correction has been made for this effect.

Table I

Descriptive Statistics on Difficulty and Reliability of

SLEP Item Types

(Based on 310 Foreign Students Entering Grades 7 through 11.)

Item Type	Listening Dictation	Cartoon	Single Picture	4-Picture Options	Map	Extended Conversation	Cloze	Literary Passage
Section	LC	RC	LC	RC	LC	LC	RC	RC
Number of Items	20	12	25	15	12	18	40	8
Mean Percent Correct	90	89	85	78	74	68	62	48
Reliability*	.889	.742	.811	.707	.721	.840	.891	.685
% Usable Pre-tested Items	65	79	75	75	86	88	93	88

*Kuder-Richardson Formula 20.

Table II

SLEP Descriptive Statistics by Section

(N = 310)

Section	Listening Comprehension	Reading Comprehension	Total Test
Number of Items	75	75	150
Reliability*	.94	.93	.96
Mean Percent Correct	81	68	74
Mean Scaled Score	25	25	50
Mean R-Biserial of Items With Section Score	.61	.55	

*Kuder-Richardson Formula 20.

Table III

Intercorrelations of SLEP Parts and Sections

	Part	1	2	3	5	6	7	Section
1.	One Picture							.84
2.	Dictation	.67						.78
3.	Map	.64	.55					.75
4.	Conversation	.72	.55	.66				.82
5.	Cartoon							.61
6.	Four Pictures				.56			.80
7.	Cloze				.53	.77		.90
8.	Literary Passage				.36	.56	.62	.67

Table IV

SLEP Mean Scores by Instructional Program

Program*	N	LC	RC	Total	S.D.
Bilingual Education, Full-Time	71	15.1	16.5	31.6	8.6
Bilingual Education, Part-Time	85	17.9	19.0	36.9	11.2
ESL, Full-Time	159	18.5	18.9	37.5	9.6
ESL, Part-Time	694	21.5	21.4	42.9	11.0
Regular	211	25.3	25.1	50.4	12.1

*No response = 19.

Table V

Correlations Between SLEP Scores and Five Demographic Variables*

	N	Listening	Reading	Total
Years English Study	1138	.40	.38	.41
(Within US)	1112	.37	.27	.34
(Outside US)	992	.13	.12	.13
Time in this School	1215	.27	.21	.25
Time in US	1220	.37	.29	.35

*All correlations are significant at the $P < .0001$ level or less.

REFERENCES

American Psychological Association. 1974. Standards for Educational and Psychological Tests. Washington, D.C.: American Psychological Association.

Angell, Ann. 1980. Test Analysis: Secondary Level English Proficiency Test. SR-80-83 Unpublished Statistical Report. Princeton, NJ: Educational Testing Service.

DiFiore, J.V. 1980. Specifications for a Listening-Dictation Item-Type. University of Florida, Unpublished Master's Thesis.

Educational Testing Service. 1979. SLEP Supervisor's Manual. Princeton, NJ: Educational Testing Service.

Educational Testing Service. 1980. Secondary Level English Proficiency Test. Princeton, NJ: Educational Testing Service.

Educational Testing Service. 1981. Manual for Administering SLEP. Princeton, NJ: Educational Testing Service.

Lado, Robert. 1961. Language Testing. New York: McGraw Hill.

SECONDARY LEVEL ENGLISH PROFICIENCY TEST

Student Questionnaire

Directions: First, print your name and write your birthdate on the lines below exactly as you did on your answer sheet. Then, before you answer each question, read the question carefully. Mark the space next to the answer you choose. Mark only one answer for each question.

NAME: _____ BIRTHDATE: ___/___/___
(Family Name) (First Name) (Middle Initial) (Month) (Day) (Year)

1. What is your current student status?
   a. [__] Foreign Student
   b. [__] Immigrant
   c. [__] Refugee
   d. [__] U.S. Citizen (Born in United States)
   e. [__] Non-Documented

2. In which of the following programs are you currently enrolled?
   a. [__] ESOL (English to Speakers of Other Languages) Program, Full Time
   b. [__] ESOL (English to Speakers of Other Languages) Program, Part Time
   c. [__] Bilingual Program, Full Time
   d. [__] Bilingual Program, Part Time
   e. [__] Regular Class (with native English speakers)

3. How long have you been enrolled at this school?
   a. [__] Less than 6 months
   b. [__] 6 months to 1 year
   c. [__] More than 1 year

4. How long have you lived in the United States?
   a. [__] Less than 6 months
   b. [__] 6 months to 1 year
   c. [__] More than 1 year, but less than 2 years
   d. [__] More than 2 years, but less than 5 years
   e. [__] More than 5 years, but not all my life
   f. [__] All my life

5. How long have you studied English in the United States?
   a. [__] Less than 1 year
   b. [__] More than 1 year, but less than 2 years
   c. [__] More than 2 years

6. How long have you studied English outside the United States?
   a. [__] Less than 1 year
   b. [__] More than 1 year, but less than 2 years
   c. [__] More than 2 years

Your answers to these questions will not be used to determine your scores on the Secondary Level English Proficiency Test. Also, your answers will not be given to anyone at your school or any other school. The information you provide by answering these questions will be used for research studies and no individually identifiable records will be maintained.